10-801: Advanced Optimization and Randomized Methods

Homework 2: Subdifferentials, SDP relaxations

(Feb 5, 2014)

Instructor: *Suvrit Sra* ***Due:*** *Feb 11, 2014*

Visit: http://www.cs.cmu.edu/~suvrit/teach/ for academic rules for homeworks.

1. Let $C \subset \mathbb{R}^n$ be closed convex set. Consider the function

$$d_C(x) := \inf_{y \in C} \|x - y\|,$$

where $\|\cdot\|$ is the Euclidean norm.

   (a) Prove that d_C is convex

   (b) For $x \in C$, what is $\partial d_C(x)$?

   (c) If $x \notin C$, prove that d_C is differentiable and show that

$$\nabla d_C(x) = \frac{x - x^*}{\|x - x^*\|},$$

   where x^* is the nearest point to x in C.

2. Show how to compute *one* subgradient for the following convex functions:

   (a) $f : \mathbb{R}^n \to \mathbb{R} \equiv x \mapsto \sum_{i=1}^{G} \|x^i\|_\infty$, where x is partitioned into subvectors as $x = [x^1, \ldots, x^G]$

   (b) $f(X) = \sum_i \sigma_i(X)$, where $\sigma_i(X)$ denotes the i-th singular value of a matrix X. (*Hint:* First show that $\sum_i \sigma_i(X) = \max_{\|Y\|_2 \le 1} \operatorname{tr}(X^T Y)$ where $\|\cdot\|_2$ is the operator norm for matrices.)

   (c) $f(x) = \lambda_{max}(e^{-\sum_{i=1}^n A_i x_i})$, where λ_{max} is the maximum eigenvalue and A_i are fixed $n \times n$ matrices.

3. Let C be an $n \times n$ symmetric positive definite matrix. Let s and y be vectors in $\mathbb{R}^n$ such that $s^T y > 0$. Consider the optimization problem

$$\inf\{\operatorname{tr}(CX) - \log\det(X) \mid Xs = y,\ X \succ 0\}.$$

   (a) Show that this problem admits a global minimizer.

   (b) Prove that for the above problem, the point

$$X = \frac{(y - \delta s)(y - \delta s)^T}{s^T(y - \delta s)} + \delta I$$

   is feasible for small $\delta > 0$

   (c) Use first order (constrained) optimality conditions to find the solution. (*Note:* This solution furnishes one of the most important substeps in nonlinear programming algorithms.)

4. SDP relaxations

   (a) Consider the 2-SAT problem for a collection of binary variables with constraints on pairs of variables. Provide a procedure to turn a 2-SAT problem into SDP form if this transformation is possible.

   (b) Express k-way clustering as a maximum cut problem – maximize the sum of distances for labels not in the same cluster. Now relax this problem into an SDP.

5. Explain why Shingling (selecting the bottom-k entries) is not exactly equivalent to the Jaccard coefficient.

6. [*Bonus*] Analyze the explicit runtime for reduplication with shingles.